The Other 50% of Internet Traffic

Craig Labovitz
labovit@deepfield.net

DeepField
NETWORKS

Earlier

- In previous work, focused on HyperGiants

  - 50% of Internet traffic due to 150 companies

  - i.e. Google, MSFT, CDNs, large consumer

- This talk looks at the other 50%...

DeepField
NETWORKS

Mysteries

- Earlier work raised number of questions

- Easy to understand role of Google, Facebook, etc.

- But harder to understand role of a handful of companies that kept showing up in the data

DeepField
NETWORKS

Mysteries

- Same ASN show up again and again

    AS16265   AS20473 $\boxed{\text{and more}}$

      AS16276   AS39572

- Same CIDR ranges and, often, same data centers (or at least traceroute paths)

- Massive share of global Internet traffic

- Asking around, many theories, but no one completely sure

DeepField
NETWORKS

Mysteries

- Same ASN show up again and again

  AS16265 AS20473   and more

  AS16276   AS39572

- Same CIDR ranges and, often, same data centers (or at least traceroute paths)

- Massive share of global Internet traffic

- Asking around, many theories, but no one completely sure (or at least had any data)

DeepField
NETWORKS

Mysteries

- So, where are these terabytes of traffic going?

- Services usually behind URL load balancer

- Netflow provides no clues...

- Other tools come up empty (whois, curl, DNS)

- Intentional service obfuscation (e.g. short lived DNS)

- Sometimes mocking

403 Forbidden

nginx/1.0.3

**You are too stupid.
This site is not for you.**

Some Words about the Data

- Different data than earlier work

  - Population discovery algorithms combined with passive / active DNS, large scale crawling, flow, third-party public / licensed data sources, some machine learning and a bit of cleverness

  - Plus large numbers of VMs around the word

- And a few months of big data crunching

DeepField
NETWORKS

About this Work

- This work a best-effort (really a side / spare-time research project)

- Data is still preliminary, but is reasonably representative and covers very large datasets...

- And finally, focus of work is on the infrastructure and economics (not the content, morality, legality, etc.)

- Complete report / research paper in progress

DeepField NETWORKS

Expectations

- Conventional wisdom* is that adult sites, P2P (trackers and seed) and file sharing are distributed across a huge swath of the Internet. Basically everywhere.

* as explained by Slashdot

DeepField
NETWORKS

Expectations

- Conventional wisdom* is that adult sites, P2P (trackers and seed) and file sharing are distributed across a huge swath of the Internet. Basically everywhere.

- No.

DeepField
NETWORKS

This Talk

- A brief tour of a small number of companies that play a massive role in Internet traffic

- Specifically, File Sharing, P2P, and Adult

- Particular focus on how this infrastructure is built and the business models that support it

DeepField
NETWORKS

File Sharing

- First have to define what we mean

- All file storage sites (well, except MegaUpload) look identical...

  - Similar graphics and sales messaging

  - Similar DMCA notices and warnings

  - So we define "file sharing" sites by industry self-classification

DeepField NETWORKS

Industry Self-Classification
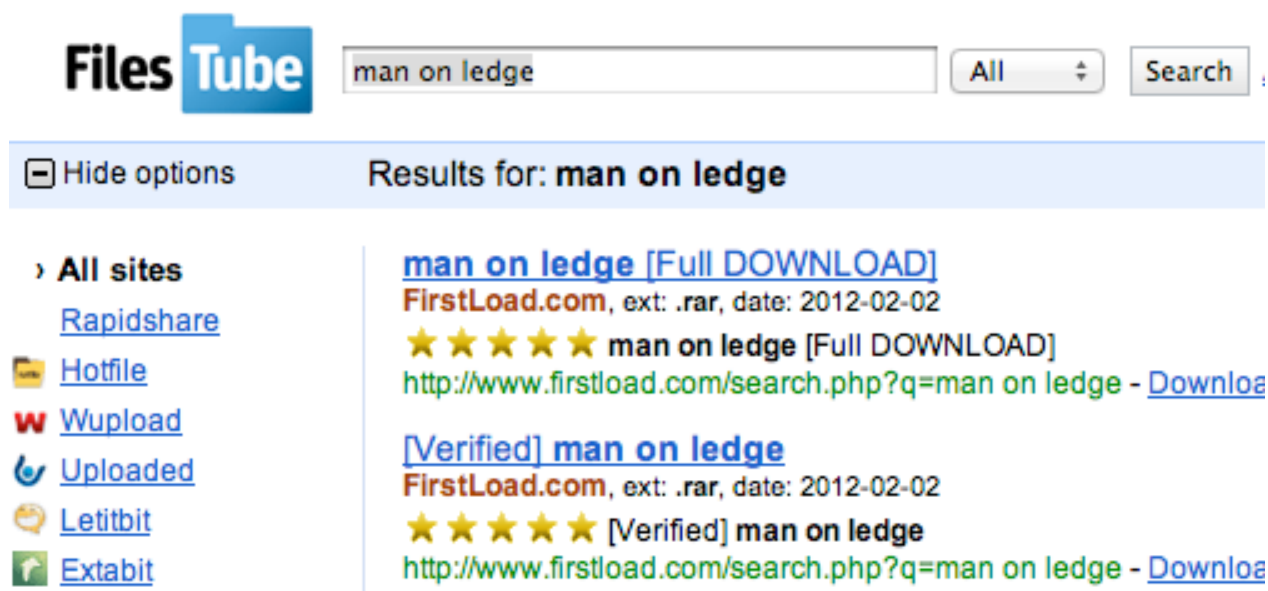
- Many dozens of "pay-for-link" sites offer the same set of file locker sites

- And generally (never) do not include dropbox, box.net, etc.

File Sharing

Distinct ecosystems each with unique infrastructure, advertising, payment and analytics partners

Multiple different business models and niche target markets

Pay for search   ($10 / month)
Pay for storage ($10 / month)
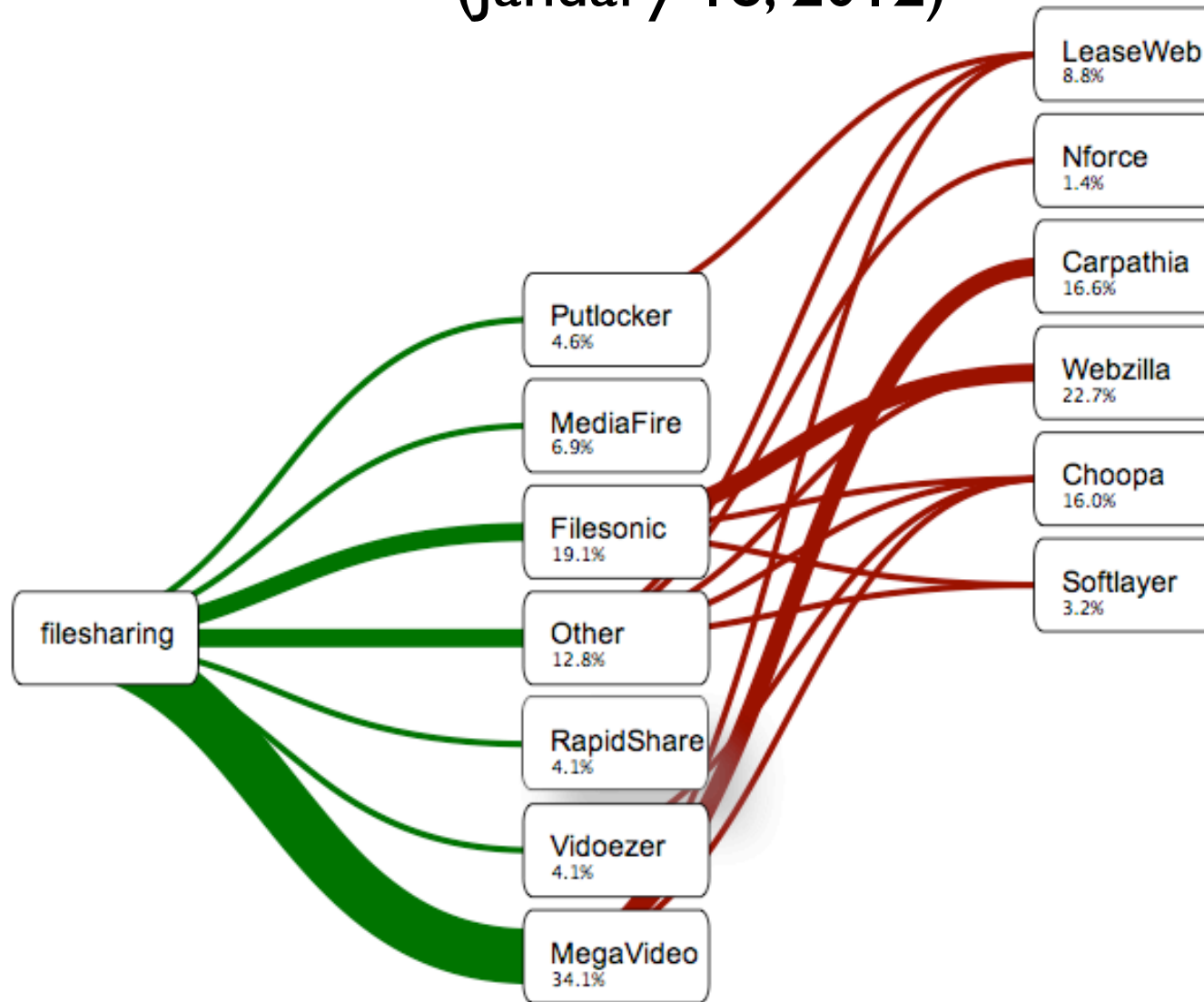
Advertisement supported

File Sharing

- Averages 5-10% of all consumer traffic

- And 1-2% of all consumers

  Peak percentage of active subscribers using file sharing service in a one hour period

- Similar traffic patterns to P2P
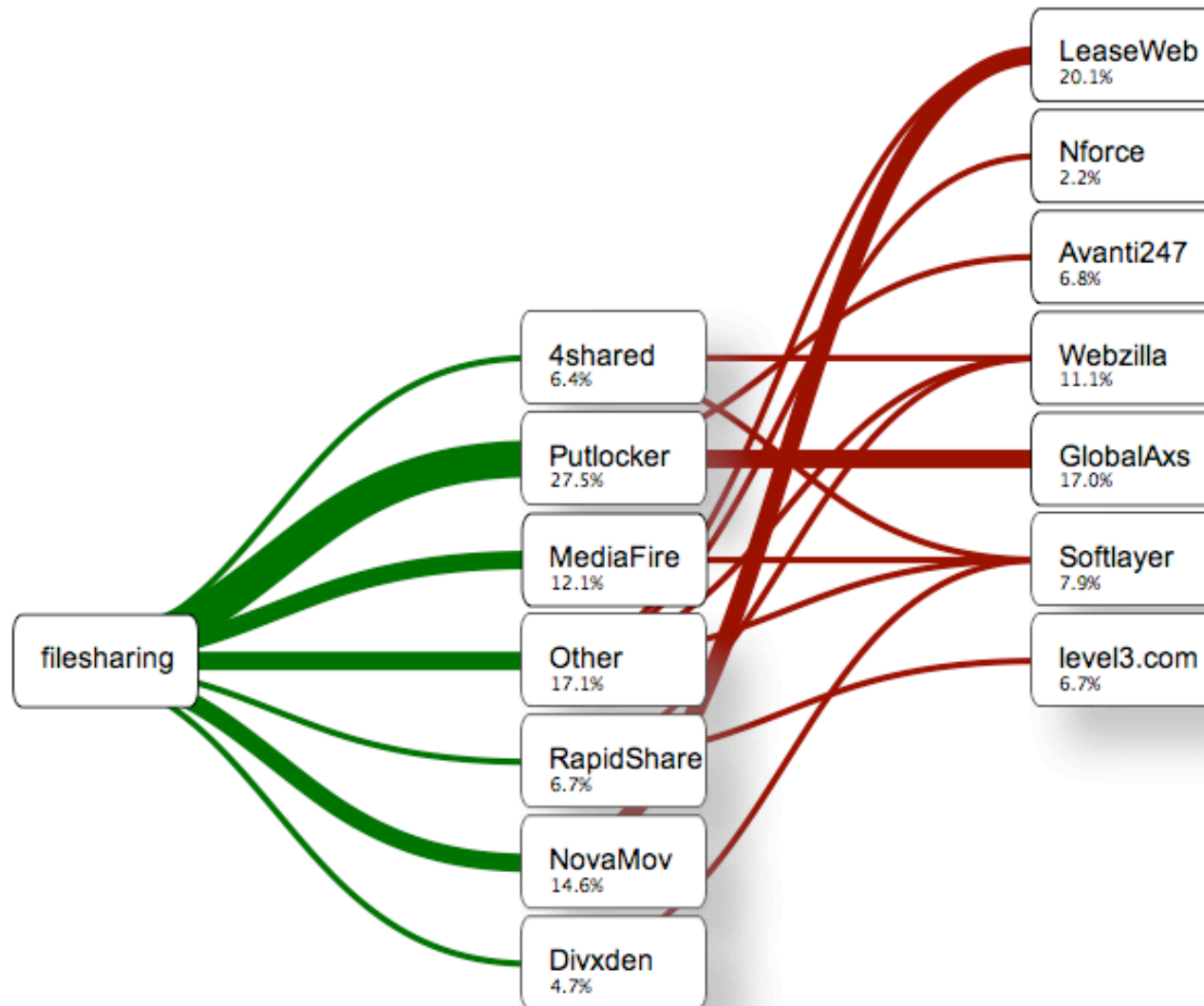
- i.e. highly diurnal and peaks around 1am local time

DeepField
NETWORKS

FileSharing

(January 18, 2012)

Lines are proportional to traffic percentages

FileSharing

(January 19, 2012)

Lines are proportional to traffic percentages

File Sharing Topology

- Hundreds of "distinct" domain names (though many actually the same site and owned by same organization)

- 10 of these sites contribute 70% of all Internet file sharing traffic

- And 4 colo / hosting companies (across ~8 locations) contribute 85% of all file sharing traffic

- Megavideo (Carpathia / Leaseweb) quickly replaced (with Putlocker a big winner)

DeepField
NETWORKS

Adult

- DRM / billing / control uses specialized hosting

- Almost traffic comes from CDNs

  - Adult sites share "red-light" neighborhood

  - Hosting similarly segment infrastructure

- Several small CDNs specialize in only adult content (e.g. Swift)

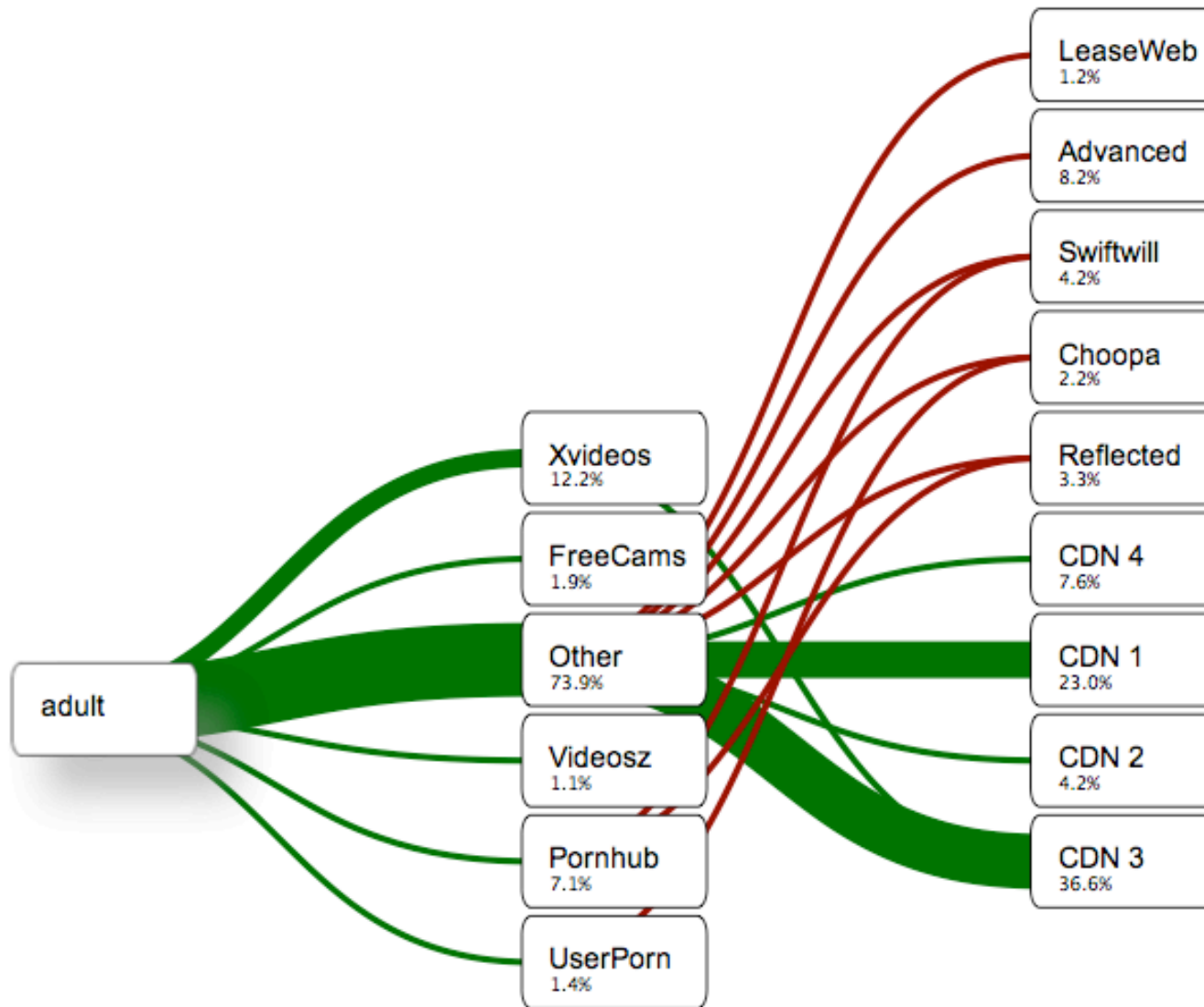- Some CDNs have decided porn is not part of their business (e.g. Akamai)

Adult

- Others cater to the market with specialized pricing, SLAs and marketing

    This means we do not ask what you stream nor do we judge what you stream, how you stream or when you stream.

- Specialized hosting, payment and advertising

Internet Adult Traffic

LeaseWeb
1.2%

Advanced
8.2%

Swiftwill
4.2%

Choopa
2.2%

Reflected
3.3%

CDN 4
7.6%

CDN 1
23.0%

CDN 2
4.2%

CDN 3
36.6%

Xvideos
12.2%

FreeCams
1.9%

Other
73.9%

Videosz
1.1%

Pornhub
7.1%

UserPorn
1.4%

adult

Lines are proportional to traffic percentages

DeepField
NETWORKS

The Rise and Fall (and Rise) of P2P

- P2P used to be mainly between dorm and consumer PCs

    - Slow, unreliable, throttled by ISPs

    - And home IPs pursued by RIAA agents

- So P2P declined as a percentage of Internet traffic (see my last talks).......

DeepField
NETWORKS
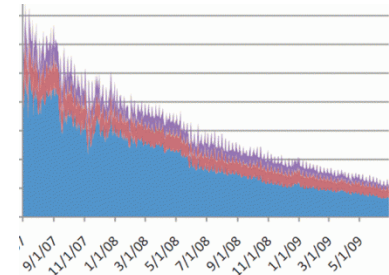
P2P Today

- But then came the cloud (and HD movies)

- Enter hosted cloud seed boxes

  - GigE interfaces (great ratios)

  - No throttling

  - Latest torrent code pre-installed

  - Conveniently located outside US (and RIAA) jurisdiction

- Still small percentage of P2P (1-5%), but growing rapidly

Extreme Seed
Seedbox hosting at its best!

Large ruTorrent Seedbox
Free Setup Within 24 Hours

» 250 GB Disk Space
» 1Gbps Unlimited Traffic   **$25**
» Unlimited Active Torrents   /month
» FTP & HTTP Downloads
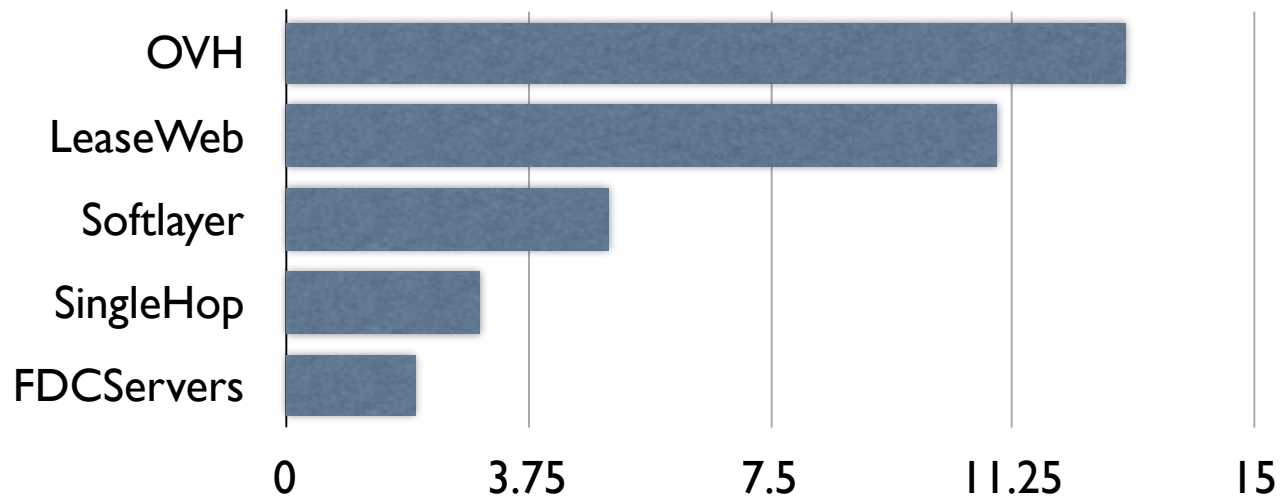» ruTorrent 3.3

order now

■ Dedi **Seedbox**

DeepField
NETWORKS

P2P Infrastructure

Seedbox / Tracker Hosting Percentages

This is a no-brainer

Softlayer's network has much better quality and peering to everywhere in the world.

OVH is probably the worst of the 3 top datacenters in the world, that is why their prices are so cheap..

Softlayer>LeaseWeb>OVH

http://torrent-invites.com/seedbox-discussions/182440-ovh-vs-softlayer.html

DeepField
NETWORKS

A Very Brief Tour

Large Russian File Sharing Hosting Provider

The new data center "Bunker"

- Colocation in the "Bunker"
- A separate room in the "Bunker"
- Live video
- The big picture

In the past few years our data center on the street. Turgenev, 52-58 as well as communication platforms for Leontovich, 9, and Gaidar, 50 have ceased to meet the growing needs of our customers, and we had a need to build a new data center.

In early 2010, the problem goes away the best - we got wasted antiradiation seekers former electromechanical plant and substation canned with enough headroom for us and connect to two independent 10-kilovolt routes.

DeepField NETWORKS

A Very Brief Tour

Large UK File Sharing Hosting Provider

Work Begins On New Datacentre

With our original datacentre already close to full capacity in Manchester we are in the process of opening our second UK datacentre right next door to our Manchester head office. This new building presents us with a white sheet of paper on which to design a state of the art datacentre which is likely to be the most energy efficient in Europe.

Final Thoughts

- File Sharing, P2P and Adult consume large proportion of Internet traffic (up to 30%)

- Tendency towards centralized / common infrastructure

  - network effect

  - specialized market needs

  - targeted sales strategy

- P2P and file sharing staggeringly inefficient (US -> Europe)

- Dynamic market with shifting demands and potentially significant impact on carrier peering / IXP decisions

DeepField
NETWORKS

Questions?

labovit@deepfield.net

DeepField
NETWORKS